Three-Dimensional Hand Tracking Based on Microstructure of Hand Motion

Fanwen Min
School of Information Science and Engineering, University of Jinan, Jinan, China
Shandong Provincial Key Laboratory of Network based Intelligent Computing, Jinan, China
Email: sdtzmfw@gmail.com

Zhiquan Feng*, Yuanyuan Su, Tingfang Zhang
School of Information Science and Engineering, University of Jinan, Jinan, China
Shandong Provincial Key Laboratory of Network based Intelligent Computing, Jinan, China
Email: fzqwww@263.net

***Abstract*— At present, hand tracking based on particle filter obtains current 3D hand gesture mainly by two elements, that is, by current frame of gesture images and by status information of the previous 3D hand gesture. Compared with current algorithms, the proposed algorithm focuses on behavior analysis and process modeling. Focusing on mathematical linkages establishment between microstructure model and the current frame gesture, the status of current 3D hand gesture is obtained not only on the basis of local frame information, but gesture model of overall movement process. After that, a unified and efficient data structure is presented, upon which the particle number can be reduced. First, hand motion law in specific circumstance is obtained by experiments based on data glove in the three-dimensional virtual platform. Second, the law is summarized as the microstructure of hand motion. Hand motion model is put forward and applied to forecast period of the particle filter algorithm, upon which hand tracking can be guided by the summarized microstructure. At last, hand tracking algorithm is carried out by a camera-oriented virtual three-dimensional interactive platform. The experimental result shows that, compared with Gaussian Process Annealed Particle Filter (GPAPF) gesture tracking algorithm, the time cost of the proposed algorithm was nearly 29% lower, while the accuracy 26% higher.**

***Index Terms*— 3D freehand tracking; particle filter; microstructure of hand motion; cognitive-behavioral model**

I. Introduction

With the emergence of new technology and new equipment, the traditional human-computer interaction is more and more difficult to meet the information exchange between electronic devices and people, new interactive ways have attracted much more attention. In new interactive approach, where "people" is the center, achieving natural, harmonious, convenient interaction is the goal. Hand tracking and interaction has been widely concerned, and three-dimensional hand tracking has become a very challenging research hotspot. Effective, real-time and accurate tracking results play a significant role in the promotion of motion capture, gesture recognition and human-computer interaction [1-3].

Currently, three-dimensional hand tracking technology can be divided into two categories [4]. One needs wearing auxiliary equipment, such as data glove and marks made on key parts of hands. Such techniques require auxiliary equipment [5], which limits the natural movement of the hand. B. Thomas et al. developed a system that used data glove as input information to control remote targets. In addition, Robert Y. Wang et al. [6] carried out real-time three-dimensional hand tracking through a colored glove and a camera. The approach in [6] was to print specific pattern on the glove, and then used gesture database and sampling search mechanism to accomplish gesture tracking. The other tracking technology approach is based on vision. With camera as the only input equipment, this kind of technology needs no more auxiliary equipment and marks on hands; it can achieve a more natural, convenient human-computer interaction.

Vision-based hand tracking method also includes two categories: one is appearance-based [7]; the other is the model-based [8]. The appearance-based tracking method does not require initialization and the tracking speed is fast. This method usually directly identifies the hand model state through the input image. This method gets a lot of image samples, and then uses statistical learning method to find out the relation between the image samples characters and the hand model. Finally this method identifies the state of hand model through regression function or classifier. Since the number of training samples can affect the accuracy of regression function or classification, a large number of dense samples need to be trained, which is the major drawback of this method. Compared with the appearance-based method, model-based approach needs to be initialized. But in the tracking process a variety of hand constraints, such as static constraints and dynamic constraints of hand joints, can be integrated into the model-based approach. This method needs to project the three-dimensional hand

doi:10.4304/jsw.9.8.2133-2141

model onto the 2D image, and then evaluates the similarity between the obtained 2D projection image and the input image. Finally this method gets the best hand state estimation from state space using optimization methods. However, as human hand is high-dimensional, model-based method requires searching the best hand state estimation in a highly dimensional space, so the efficiency is relatively low; it is the main drawback of this method [9].

As a solution to the non-Gaussian and nonlinear problems developed from the late 1990s, particle filter has been considered to be one of the most promising method to estimate state. Non-Gaussian and nonlinear problems caused by occlusion and background complexity can be dealt with effectively by the method. In each frame of tracking, the output of this method is the probability approximation of the tracking target location, and the method can retain multiple estimates spreading over time. With this feature, the method can determine the target and continue to track target when the target is lost. Besides, particle filter has strong robustness in dealing with hand occlusion, complex background changes, illumination changes and other factors that affect the tracking [10].

Given that particle filter has a good feature in solving tracking problem, it has become one of the main technical means of hand motion tracking. However, for a high-dimensional space of human motion tracking, the particle filter requires a lot of samples to simulate the posterior distribution [11]. Therefore, in order to achieve high estimation accuracy, particle filter takes large amount of computation [12]. To solve the problem, researchers have conducted a lot of studies. The current practice is to reduce the dimension of the state space, so that the number of particles in the process of sampling is reduced, thus computation reduced.

As a model-based approach, analysis - synthesis technique [13, 14] is a widely used in the methods of reducing the dimension of the state space. First, feature detection and parameter estimation are needed for the 3D hand model. Then based on the prior knowledge and hand motion constraints, the model parameters are estimated by stepwise optimization. Finally, the approving 3D hand state estimation is obtained. The disadvantage of this method is that the time cost of stepwise optimization is too large, and it also lacks of optimization function evaluation criteria.

Unlike the analysis - synthesis technique, the method of state space partition is relatively straightforward. Here is how it works: dividing the high-dimensional state space into a number of subspaces firstly, then using particle filter tracking separately for each subspace; or, dealing with a separate subspace firstly, then using the state estimate obtained as a constraint to deal with the remaining subspace. This method has the advantage of higher tracking precision, robustness and real-time performance. But reasonable division of space is difficult. Chang et al. [15] used the hierarchical search method, where high dimensional space was decomposed into several low dimension spaces, thus reducing the time cost due to high degrees of freedom.

Another effective way to solve the problem of high-dimensional hand space is to establish the relationship between human hand and high dimensional feature set. The relationship can be obtained by deformation technology, popular learning and artificial neural networks, and so on. Using a three-dimensional graph model, Tongli Liu et al. [2] described hand structure, dynamics, kinematics and self-occlusion. The problem of tracking a 27-dimensional hand model was transformed into tracking 16 smaller problems; each of these problems was a 6-dementional model. All the 16 smaller problems were handled at same time, thus reducing the computational complexity in the process of hand tracking. Guan et al. [16] represented the samples of the high-dimensional space using ISOSOM (isometric self-organizing map) low-dimensional popular space, during which process the high-dimensional space was hierarchical divided. Then the context of hand feature was obtained from the input images, finally H - ISOSOM algorithm was used to estimate the state of hand.

Different hand tracking methods have emerged these years, yet they only solve the problem under certain conditions. There is still a big gap between natural, convenient and harmonious human-computer interaction and the existing tracking methods. Therefore, researchers have begun to turn gaze to the "people" and cognitive theory is introduced in the study of hand tracking. Zhiquan Feng et al. [17] provided a new perspective for natural hand tracking and interactive algorithm. They analyzed the cognitive process and behavior in the process of gesture interaction and, by doing so, revealed and elaborated the interactive laws and motion characteristics.

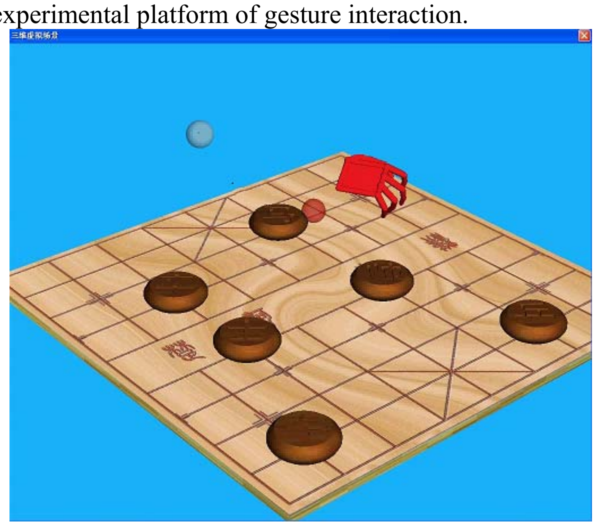
Current particle-filtering-based hand tracking obtains current 3D hand gesture by current frame gesture image and status information of the previous 3D hand gesture. This approach focuses on the analysis of local information but ignores the overall process, so it may have lost continuity information of gesture motion. Compared with the existing methods, this paper presents the particle filter algorithm based on microstructure of hand motion. With behavior analysis and process modeling as the entry point, this method focuses on mathematical linkages establishment between microstructure model and the current frame gesture. The status of current 3D hand gesture is obtained on the basis of gesture model of overall movement process. Microstructure is used to guide hand tracking, which guarantees the consideration of continuity information of hand motion and local information of the current frame. It provides a unified and efficient data structure and sampling methods for particle formation process. This reduces the particle number of sampling, thereby the time cost of hand tracking is reduced and the real-time performance is improved.

II. ESTABLISHMENT OF MICROSTRUCTURE OF HAND MOTION

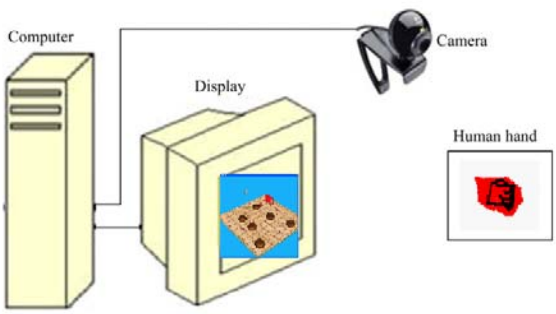
A. *The Experimental Platform of Gesture Interaction*

In order to analyze the natural hand tracking and 3D interaction, two three-dimensional interactive experimental platforms with the same scene and different input devices have been built: experimental platform based on data glove and experimental platform based on camera. The complete experimental process is to interact in the three-dimensional virtual scene, which is constructed with a Chinese chess board and six pieces. Six pieces on the chessboard need to be translated from the initial position to the final position.

3D interactive platform based on data glove mainly uses data glove (5DT Data Glove 14 Ultra) and position tracker to obtain the real-time data in the interactive process. Operators use data glove and position tracker as input devices to control hand translation, rotation, grasping and releasing and other actions. By analyzing the data generated in these processes, the law which provides the direct support for the establishment of microstructure can be found. Camera-based virtual three-dimensional interactive platform is mainly used to verify the proposed algorithm. Through the microstructure establishment of hand motion, the microstructure is applied to the 3D hand interactive process. The hand tracking algorithm is proposed on the basis of microstructure, and then carried out through the establishment of camera-based platform. Fig. 1 shows the experimental platform of gesture interaction.



(a) The 3D virtual scene



(b) The hardware platform of system, only using one camera to get video input

Figure 1. The experimental platform of gesture interaction

B. *The Behavioral Modeling of Gesture Interaction*

By analyzing the experimental data got by data glove and position tracker, the interactive experimental processes can be divided into four stages. Reference [17] described the four stages: 1) the stage of gesture translation; 2) the stage of grasping an object; 3) the stage of gesture transition with an object in the hand; 4) the stage of releasing the object. Fig. 2 shows the four stages and the hand status diagrams of each stage. The hand characteristics of the four stages are analyzed and summarized, and then laws in each stage of hand motion can be obtained. These laws are summarized as microstructure of hand motion, and then hand tracking of each stage can be guided through the microstructure.

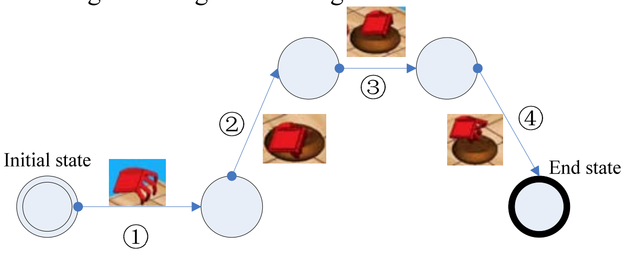


Figure 2. The four behavioral stages of gesture interaction

C. *Hand Motion Model of Grasping and Releasing*

The tracking of finger joints is the key of 3D hand tracking. In the four stages of interactive experiment, the hand gesture of stage 1) and stage 3) is in the process of translation, where the finger joints are hardly changed, so it is particularly important to model the finger joints motion of stage 2) and stage 4). The 3D hand model used in this paper was proposed by Xu [18]. Data of hand joints changes in interactive experiments is obtained through data glove. The experiment for PIP (the joint between middle and proximal) on the index finger is repeated 10 times, and the result curves are shown in Fig. 3.

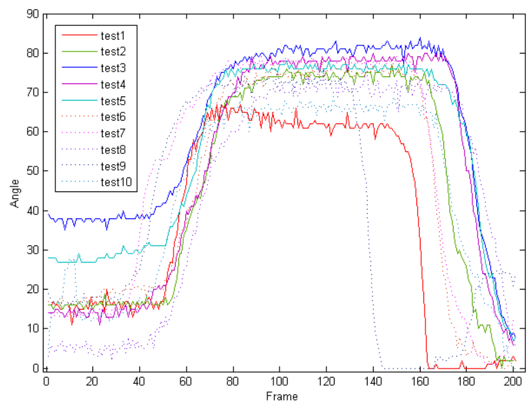


Figure 3. The parameter changes of the PIP on the index finger in 10 repeated experiments

The curves in Fig. 3 show that the changes of mathematical curves have consistency in the same interactive experimental task. That is to say: the finger joints motion have certain regularity. By fitting the curves of grasping and releasing by MatLab, hand motion prediction model of grasping and releasing can be achieved.

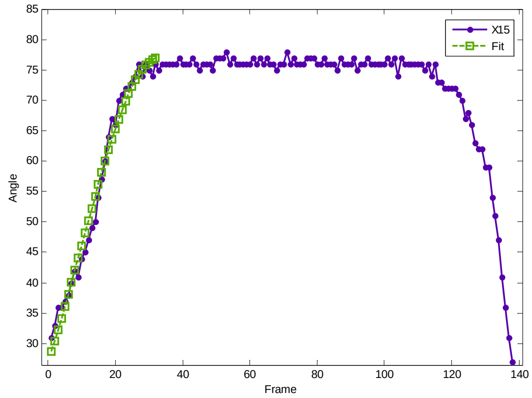


Figure 4. The grasping motion fitting of the PIP on the index finger

Fig. 4 shows the Gaussian fitting of the PIP on the index finger when hand gesture is in the process of grasping. It can be known from the fig. 4 that Gaussian fitting can better express hand joints motion of grasping, and hand joints motion of releasing is equivalent to the inverse process of grasping. The fitting formula is

$$X^k(x) = a_1 e^{-(\frac{x-b_1}{c_1})^2} + a_2 e^{-(\frac{x-b_2}{c_2})^2} \qquad (1)$$

In the above formula, a_1, a_2, b_1, b_2, c_1, c_2 are parameters of the Gaussian fitting; k is the k-th degree of freedom of the 3D hand model, and the range of k is 10, 11, 12, 14, 15, 16, 18, 19, 20, 22, 23, 24, 26, 27, 28; $k \in (10,11,12)$ denotes the angle of the three joints on the thumb; $k \in (14,15,16)$ denotes the angle of the three joints on the index finger; $k \in (18,19,20)$ denotes the angle of the three joints on the middle finger; $k \in (22,23,24)$ denotes the angle of the three joints on the ring finger; $k \in (26,27,28)$ denotes the angle of the three joints on the pinkie. That is, (1) exactly represents the angle changes of all hand joints.

III. The 3D Hand Tracking Based on Microstructure of Hand Motion

When designing the PF (Particle Filter) [19, 20] algorithm, it is necessary to consider the following aspects: First, how to use fewer sample of gesture state to truly reflect the gesture distribution; Second, how to reduce the time cost of hand tracking, given that the accuracy of hand tracking should be ensured. For different stages, gestures can be processed respectively. Therefore, a complete gesture interactive process can be divided into several stages where time series are connected but state changes are different. For different parts of hand state, the mutative degrees of freedom and constant degrees of freedom in 3D hand model can be distinguished. The purpose of dimension reduction can be achieved by simply keeping tracking the mutative degrees of freedom. At the same time, the real distribution of gesture can be better reflected according to the hand motion model in the specific environment, which guarantees the accuracy of hand tracking.

Based on the above ideas, the stage of hand tracking will be predicted after the object selection. Different stages will be processed differently. In the prediction phase of PF, the microstructure of hand motion is used to guide the hand tracking. The microstructure of hand motion is not only beneficial to design program, but it is a new way of hand tracking. It is more important that the microstructure provides a unified and efficient data structure and sampling methods for the particle formation process and reduces the number of samples. Fig. 5 shows the process of the proposed three-dimensional hand tracking algorithm. Fig. 6 depicts the real-time tracking process of the "grasping" and "releasing" stage.

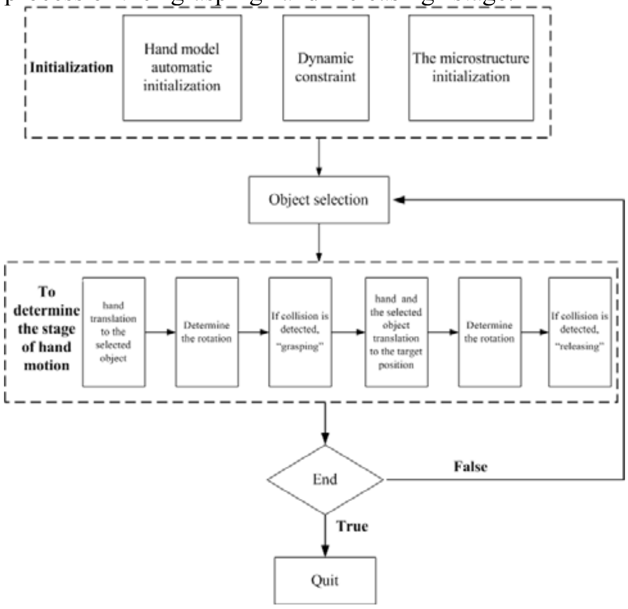


Figure 5. The process of the proposed three-dimensional hand tracking algorithm

When hand gesture is in the process of grasping and releasing, the algorithm is as follows.

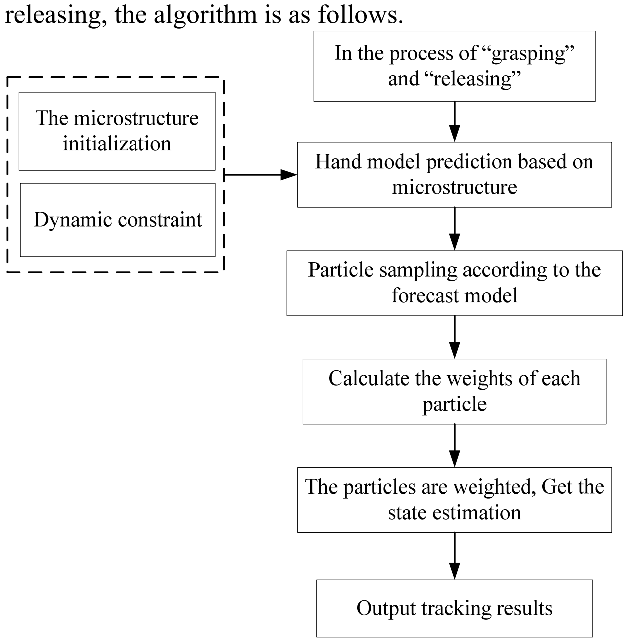


Figure 6. The real-time tracking process of the "grasping" and "releasing" stage

Specific steps of the proposed algorithm are as follows:

Step 1. Initialization.

(a) To automatically initialize Hand model. Get initial hand state according to the method proposed in [21] and [22].

(b) To determine the dynamic constraint.

(c) To obtain hand grasping model M_1, M_2, …, M_K by microstructure of hand motion.

Step 2. Object selection. The object selection process begins as soon as initialization has finished or the current object is released. In this paper the object selection is to choose the pieces that will be grasped.

Step 3. To determine the stage of hand motion. According to the behavioral model, the stage of hand motion can be predicted. The cognitive process of gesture interaction is divided into seven phases [17], while the interactive experimental process is divided into four stages. This paper has adopted these phases. According to the interactive experimental condition, the gesture rotation process is combined with the process of hand translation. A complete interactive process includes the following six stages:

(a) After selecting an object, hand begins translation from the current location to the selected object. According to the interactive experiments, the translational path adopts the curved line, while the translational speed is obtained by the moving distance between the hand centroid of two image frames.

(b) According to the interactive experiments, hand may be rotated in the process of hand translation. In the proposed system, two kinds of rotation are judged: 1) one is the rotation along the hand forward vector; 2) the other is the rotation along the vector product of the hand forward vector and the hand upward vector.

(c) Collision detection occurs when hand approaches the selected object. If collision is detected, the hand joints need to be predicted and tracked accurately. If no collision is detected, the hand continues to translate.

(d) After grasping the object, the object needs to be translated to its final position. This moment hand translates with an object in it. Translational path and the speed are processed as (a) of step 3.

(e) The rotation of hand also needs to be determined in this process of translation. At this time, hand and the object need to be rotated together.

(f) At the end of the translational path, collision needs to be detected. If any collision is detected, the object is ready to be released, and hand joints also need to be predicted and tracked accurately. If no collision is detected, the hand and the object continue to translate.

Step 4. If hand is in the phase of (c) or (f) of step 3, i.e. the hand gesture is in the process of grasping and releasing, the further prediction is made through the hand model ((c) of step 1) generated by microstructure.

{

(a) Calculate the projection sequence of each model' prior n frames $p_1[n]$, $p_2[n]$, …, $p_K[n]$;

(b) Calculate the hausdorff distance between each frame $p_i[n]$ ($1 \le i \le K$) and observed image sequence $I[n]$, Hd_1, Hd_2,…, Hd_K is obtained;

(c) Return M_i that has the minimum Hd_i.

}

The returned hand model M_i is the predicted hand model.

Step 5. For each frame of M_i, use Gaussian sampling to generate N particles. Thus gesture sample group R_j^t ($1 \le t \le N$) is obtained.

$$R_j^t = M_i^j + \eta_G \tag{2}$$

R_j^t denotes the t-th gesture sample, which is generated by the Gaussian sampling at the j-th frame. M_i^j denotes the gesture state of the i-th hand model at the j-th frame. η_G denotes the values generated by the Gaussian sampling.

Step 6. Calculate the weights of each particle ω_j^t at the j-th frame. The similarity degree between the input image and the projection of the predicted 3D hand model onto a 2D plane represents the weight of the particle. The hausdorff distance H_t between the above two images is used to measure the degree of similarity.

$$\omega_j^t = e^{-\lambda H_t} \tag{3}$$

Here, $\lambda = 0.01$; H_t denotes the hausdorff distance of the t-th gesture particle. Then the weights can be normalized.

$$\omega_j^t = \frac{\omega_j^t}{\sum_{t=1}^{N} \omega_j^t} \tag{4}$$

Step 7. State estimation. The gesture output state is got at the j-th frame.

$$R_j = \sum_{t=1}^{N} \omega_j^t R_j^t \tag{5}$$

Wherein, N represents the particle number; ω_j^t denotes the weight of the t-th gesture sample at the j-th frame.

Step 8. Loop. If not all tasks are completed, go to step 2; otherwise, exit.

IV. Experimental Results and Analysis

A. *The Experimental Equipment and Evaluation Criteria*

An ordinary USB digital camera capturing 640×480 video at 60 Hz was used. The computer used in the experiment was equipped with Inter CoreTM 2 CPU 2.66 GHz processor and 4 GB memory. The whole system had been developed in the IDE of VC++ 6.0. The particle number in the study was ten.

The evaluation criteria to verify the proposed algorithm in this paper are the average tracking accuracy and the average running time. We adopted the accuracy definition and the running time definition proposed in [4].

In a complete interactive experiment, the average tracking accuracy is defined as the average value of the accuracy of all frames. The similarity degree, which is between the input hand image and the projection of the predicted 3D hand model onto a 2D plane, represents the tracking accuracy of each frame. The distance H_d is used to measure the degree of similarity.

$$p = \exp(-k \times H_d) \tag{6}$$

In (6), k is a constant and k = 0.01. The smaller the value of H_d is, i.e., the smaller hausdorff distance between the input image and the projected image is, the more similar the two images are, while the greater the value of p is. So the greater the value of p is, the more accurate the tracking results are. The main factor affecting the average tracking accuracy is the accuracy of microstructure model. The more accurately the microstructure model represents hand and the closer it is to the model of real hand motion, the more accurately the hand model state is predicted and the higher precision the tracking is.

In a complete interactive experiment, the ratio between the total time T_{total} and the total number of frames F_{total} is called the average running time of an interactive experiment.

$$t = \frac{T_{totle}}{F_{totle}} \tag{7}$$

The smaller the average time t is, the less time cost of each frame is and the better the real-time tracking is. The total running time of this article is mainly composed of four parts: 1) the time of translation; 2) the time of rotation; 3) the time of grasping objects; 4) the time of releasing objects.

B. The Results of Accuracy and Running Time

Fig. 7 shows the experimental scene. To test the proposed algorithm in a three-dimensional virtual scene, the scene is to grasp and release the pieces on a Chinese chess board. The basic framework of the system is shown in fig. 7. The image frame that is obtained from a RGB camera is shown in the lower left corner. The upper left corner displays the segmentation of hand. The right shows the 3D hand in virtual platform. Through the real-time hand tracking, the 3D scene can be operated.

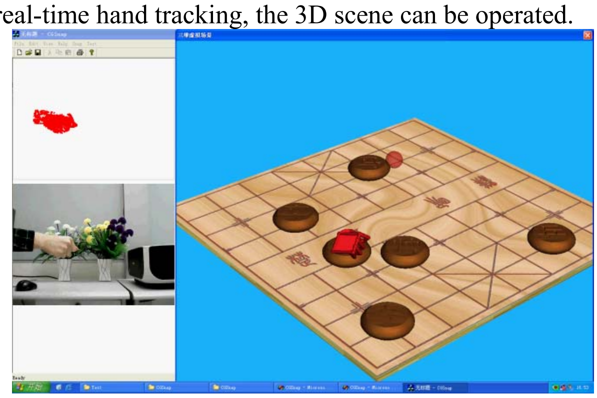
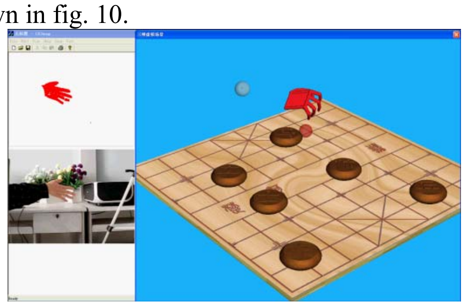
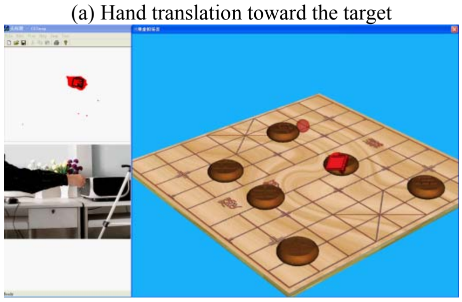


Figure 7. The basic framework of the system

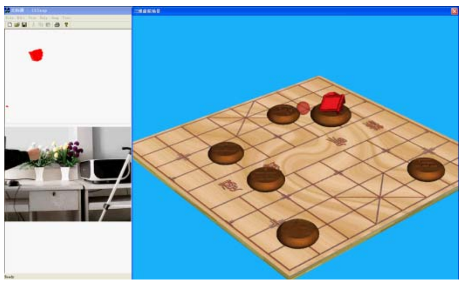
Under the conditions of a monocular camera, Gaussian process annealed particle filter algorithm (GPAPF) [23] is compared to verify the advantages of the proposed method. Operators using these two algorithms conduct experiments in the virtual platform of Fig. 7. Fig. 8 shows a process of gesture interaction: (a) Hand translation toward the target. In this process, operators choose the piece to be grasped, and then control the translation of hand in the 3D virtual scene through the translation of the real hand. In the process of hand translation, hand can be rotated. (b) Grasping the piece. If collision is detected, grasping motion occurs. At the same time, hand joints need to be predicted and tracked accurately. (c) Hand translation with an object in the hand. After grasping the piece, the piece needs to be translated to its final position. In this process, hand and the piece can be rotated together. (d) Releasing the piece. When the piece reaches its final position, collision needs to be detected. If collision is detected, the object must be released. At the same time, the hand joints also need to be predicted and tracked accurately. Fig. 9 shows the details of the operator grasping a piece: (a) is the video information captured by a camera; (b) is the hand segmentation of (a); (c) is the 3D hand model state obtained by the proposed algorithm. Based on the platform of fig. 7, the operator conducted ten times of experiment. At the same time, the accuracy and running time of each frame were recorded. The result is shown in fig. 10.



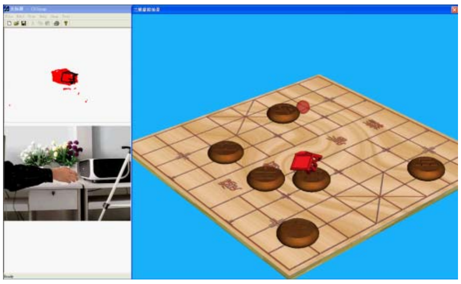
(a) Hand translation toward the target



(b) Grasping the piece



(c) Hand translation with an object in the hand

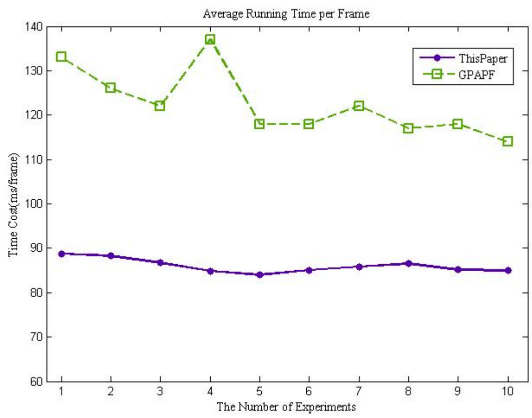


(d) Releasing the piece

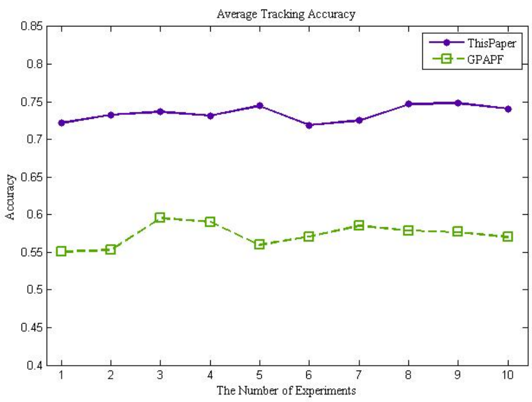
Figure 8. A process of gesture interaction

(a) (b) (c)

Figure 9. The details of the operator grasps a piece. (a) The video information, (b) Segmentation of hand and projection of 3D hand model, (c) The 3D scene



(a) The average running time per frame



(b) The average tracking accuracy

Figure 10. Average results over 10 runs of the proposed algorithm and GPAPF. (a) The average running time per frame, (b) The average tracking accuracy

As can be seen from fig. 10 (a), the time cost of the proposed algorithm fluctuates between 85 ms and 89 ms, with an average value 87 ms/frame. Whereas the average time cost of GPAPF is 122 ms/frame. The time cost of the proposed algorithm is 35 ms below, which is nearly 29% lower than that of GPAPF. This indicates that the proposed algorithm is better than GPAPF in terms of real-time tracking performance. As can be seen from fig. 10 (b), the tracking accuracy of the proposed algorithm fluctuates between 0.71 and 0.75, with an average value 0.73. Whereas the tracking accuracy of GPAPF fluctuates between 0.55 and 0.6, and the average value is 0.58. The accuracy of the proposed algorithm is 26% higher than that of GPAPF. This indicates that the proposed algorithm is better than GPAPF in terms of tracking accuracy performance. In summary, the proposed algorithm has good performance in terms of both real-time tracking and tracking accuracy. This verifies the effectiveness and superiority of the proposed algorithm.

C. The System Assessment

The time cost of this system is mainly composed of four parts: 1) the time of translation; 2) the time of rotation; 3) the time of grasping objects; 4) the time of releasing objects. To get the time cost of the four aspects in a complete experiment, the operator did 10 times of experiment. Fig. 11 shows the average time of each link. As can be seen from fig. 11, the time cost of hand translation is most, while the time cost of releasing objects is least. The most of running time is consumed in the process of translation. The proportions of the time cost of four aspects are in line with the realistic cognition. It indicates that the time cost of grasping and releasing can be accepted.

Besides the tracking accuracy and real-time performance, the 3D natural hand tracking system should also consider the cognitive burden on the operator. Evaluate the system using the proposed algorithm and the system using GPAPF from four aspects. Fig. 12 shows the four aspects: fatigue, enjoyment, convenience and feasibility; the term fatigue describes the degree of fatigue when using the system; enjoyment describes the degree of pleasure when using the system; convenience describes the convenient degree of achieving the operation purpose; feasibility describes the feasible degree of the interactive algorithm. Fig. 12 shows that, compared to the system using GPAPF, the proposed algorithm provides a 33% lower in fatigue, 56% higher in enjoyment, 13% higher in convenience, 22% higher in feasibility.

A system should also have universality. The proposed algorithm can not only be used to play chess in the virtual scene, but be used in other areas, such as virtual assembly systems. In the virtual assembly system, real-time hand tracking can be achieved by using the proposed algorithm, based on which the parts of machine can be assembled into the proper position. This virtual assembly platform has certain advantages in staff training: firstly, this assembly platform is more convenient and easy to implement than the actual platform; secondly, by using the platform to train employees, the risk is smaller and the staff can master operating steps before the actual operation. In short, the proposed algorithm can be widely used in the environment where “translation”, “grasping and releasing” and “rotation” operation are involved. With hand tracking to complete the control operation, the proposed algorithm can provide convenient, quick, casual and natural interaction.

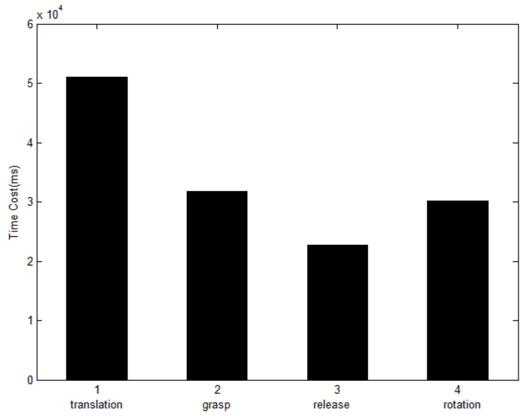


Figure 11. The average time of each link

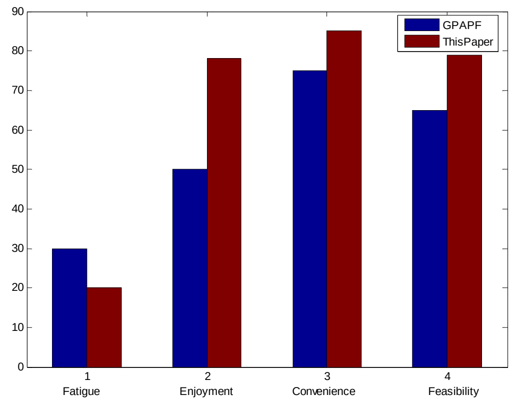


Figure 12. The feedback for users' cognitive burden

V. Conclusion and Future Work

Particle filter has strong robustness in dealing with hand occlusion, complex background changes, illumination changes and other factors that affect the tracking, so it has become a powerful tool to handle non-Gaussian, nonlinear problems and been considered to be one of the most promising method to estimate state. Current particle-filtering-based hand tracking obtains current 3D hand gesture by current frame gesture image and status information of the previous 3D hand gesture. This approach focuses on local information analysis while ignores the overall process, which may lose the continuity information of gesture motion. This leads to a high fluctuation in the particle prediction and formation.

Unlike the existing methods, the entry points of this paper are behavior analysis and process modeling. By establishing the mathematical linkages between microstructure model and the current frame gesture, the status of the current 3D hand gesture can be obtained on the basis of gesture model of overall movement process. Microstructure, which considers the continuity information of hand motion and the local information of the current frame, is used to guide hand tracking. This practice avoids the drawbacks of the existing methods and provides a unified and efficient data structure and sampling methods for the particle formation process. This reduces the particle number of sampling, thereby the time cost of hand tracking is reduced and the real-time performance is improved.

Currently, the study of three-dimensional hand tracking and interaction is still in the stage of using the particular method to solve the specific problem in a particular environment, which leads to poor portability and adaptability. Moreover, the existing methods cannot adapt to the complex and changeable actual environment [24]. In order to achieve real-time, nature 3D hand interaction, the hand interactive behavioral characteristics and hand motion variable characteristics in the specific environment are analyzed. Then the microstructure of hand motion can be summarized. Afterward, hand motion model is put forward and applied to the forecast period of the particle filter algorithm, and then hand tracking can be guided by microstructure. At last, hand gesture tracking algorithm is carried out by camera-oriented virtual three-dimensional interactive platform. The experimental result shows that, compared with GPAPF gesture tracking algorithm, the time cost of the proposed algorithm was nearly 29% lower, while the accuracy 26% higher. The future work is to improve the microstructure model. To make gesture interaction closer to the actual application, microstructure needs to be expressed more strictly, such as using mathematical language and data structures. Furthermore, we should build better hand motion models and pay attention to the natural gesture interactive problems.

Acknowledgment

This paper is supported by the National Natural Science Foundation of China (No. 61173079) and the Key Project of Natural Science Foundation of Shandong Province (ZR2011FZ003).

References

[1] Xianhui Song, Zhiquan Feng, Bo Yang, et al.. Research on Grasping Hand Gesture Based on Analysis of Occluded Information [J]. Journal of Computers, 2012, 7(3): 768-773.

[2] Tangli Liu, Xinxiao Wu, Wei Liang, Yunde Jia. 3D Articulated Hand Tracking by Nonparametric Belief Propagation on Feasible Configuration Space [J]. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(4): 476-481.

[3] Xiying Wang, Xiwen Zhang, Guozhong Dai. An Approach to Tracking Deformable Hand Gesture for Real-Time Interaction [J]. Journal of Software, 2007, 18(10): 2423-2433.

[4] Yan Lin, Zhiquan Feng, Deliang Zhu, et al.. Three-Dimensional Hand Tracking Algorithm Characterized by Multi-model Fusion [J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(4): 450-459.

[5] Yang Liu. Hand gesture interaction methods based on wearable vision [D]. Beijing: Beijing Institute of Technology, 2005.

[6] Robert Y. Wang, Jovan Popović. Real-Time Hand-Tracking with a Color Glove. ACM Transactions on Graphics (SIGGRAPH 2009), 2009, 28(3): 1-8.

[7] Athitsos V, Scaroff S. Estimating 3D hand pose from a cluttered image [C]//Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, 2003: 432-439.
[8] J Y Lin, Y Wu, T S Huang. 3D model-based hand tracking using stochastic direct search method [C] Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2004: 693-698.
[9] Zhiguo Lü, Yan Li, Xin Xu. Research on Fast 3D Hand Motion Tracking System [J]. Journal of Computer Research and Development, 2012, 49(7): 1398-1407.
[10] Zhiquan Feng, Bo Yang, Yanwei Zheng, et al.. Design and Research on Particle Filtering for Human Hand 3D Tracking [J]. Journal of Software, 2008, 19(Supplement): 87-95.
[11] Wei Gai, Zhiquan Feng, Bo Yang, et al.. 3D Hand Motion Tracking Using Improved Hidden Markov Model of Behavior[J]. Journal of Computers, 2012, 7(3): 730-735.
[12] Morshidi Malik, Tjahjadi Tardi. Gravity optimised particle filter for hand tracking [J]. Pattern Recognition, 2013, 47(1): 194-207.
[13] Rehg JM, Kanade T. Digiteyes: Vision-Based human hand tracking. Technical Report, CMU-CS-93-220, School of Computer Science, Carnegie Mellon University, 1993.
[14] Rehg JM, Kanade T. Visual tracking of high DOF articulated structures: An application to human tracking. In: Eklundh JO, ed. Proc. of the 3rd European Conf. on Computer Vision, Vol.801. Heidelberg: Springer-Verlag, 1994: 35−46.
[15] I.C. Chang, Y. Lin. 3D human motion tracking based on a progressive particle filter[J]. Pattern Recognition, 2010, 43(10):3621–3635.
[16] Guan H, Chang J S, Chen O et al. Multi-view appearance-based 3D hand pose estimation[C] //Proc of Conf on Computer Vision and Pattern Recognition Workshop(CVPRW'06), Piscataway, NJ: IEEE, 2006:154.
[17] Feng Zhiquan, Yang Bo, Li Yi, et al. Hand tracking method based on interactive behavioral analysis [J]. Computer Integrated Manufacturing Systems, 2012, 18(1): 31-39.
[18] Ting Xu. Research on Moving Hand Tracking Based on Behavior Analysis [D]. Jinan: university of Jinan, 2011.
[19] A. Doucet, S.J. Godsill, C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering [J]. Statistics and Computing, 2000, 10: 197–208.
[20] Fasheng Wang, Qingjie Zhao. A new particle filter for nonlinear filtering problems [J]. Journal of Computers, 2008, 31(2): 346-352.
[21] Zhiquan Feng, Zhang M M, Pan Z G et al.. 3D-freehand-pose initialization based on operator's cognitive behavior models [J]. Visual Computer, 2010, 26: 607-617.
[22] Zhiquan Feng, Bo Yang, Yanwei Zheng, et al.. Initialization of 3D Human Hand Model and Its Applications in Human Hand Tracking [J]. Journal of Computers, 2012, 7(2): 419-426.
[23] Raskin, Rivlin L., Rudzsky E.. Dimensionality Reduction for Articulated Body Tracking. 3DTV Conference, 2007, Kos Island, 2007, Pages 1-4.
[24] Aili Shang, Zhiquan Feng. The Application of Microscopic Structure in The 3D Hand Tracking Interactive System [J]. Journal of University of Jinan (Sci. & Tech.), 2013, 27(4): 342-346.

Fanwen Min, born in 1987, Shandong Province, China. He is an enrolled postgraduate of university of Jinan. His main research interests include human-computer interaction (HCI) and 3D hand tracking.

Zhiquan Feng, born in 1964, Ph.D., Professor, and master supervisor. His main research interests include image processing, moving human hand tracking, and human-computer interaction.

Yuanyuan Su, born in 1989, enrolled postgraduate. Her main research interest is human-computer interaction (HCI).

Tingfang Zhang, born in 1987, enrolled postgraduate. Her main research interest is human-computer interaction (HCI).